(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 176 341**
B1

(12) # EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification: **10.10.90**

(51) Int. Cl.[5]: **C 12 N 15/71**, C 12 N 15/16, C 12 N 15/18, C 12 N 15/62

(21) Application number: **85306764.3**

(22) Date of filing: **24.09.85**

(54) **An improved derivative of the tryptophan operon for expression of fused gene products.**

(30) Priority: **27.09.84 US 655183**

(43) Date of publication of application:
02.04.86 Bulletin 86/14

(45) Publication of the grant of the patent:
10.10.90 Bulletin 90/41

(84) Designated Contracting States:
BE CH DE FR GB IT LI NL SE

(56) References cited:
EP-A-0 036 776

JOURNAL OF BACTERIOLOGY, vol.133, no.3, March 1978, Washington D.C., G.F.MIOZZARI et al."Translation of the Leader Region of the Escherichia coli Tryptophan Operon" pp. 1457-1466

(73) Proprietor: **ELI LILLY AND COMPANY**
Lilly Corporate Center
Indianapolis Indiana 46285 (US)

(72) Inventor: **Belagaje, Rama M.**
7821 Mohawk Lane
Indianapolis Indiana 46260 (US)
Inventor: **Epp, Janet Kay**
920 North Graham Avenue
Indianapolis Indiana 46219 (US)
Inventor: **Hoskins, Jo Ann**
8229 Tern Court
Indianapolis Indiana 46256 (US)
Inventor: **Hsiung, Hansen Maxwell**
108 West 88th Street
Indianapolis Indiana 46260 (US)
Inventor: **Long, George Louis**
6532 Wyman Court
Indianapolis Indiana 46220 (US)
Inventor: **Schoner, Brigitte Elisabeth**
274 Larkspur Court
Zionsville Indiana 46077 (US)

(74) Representative: **Hudson, Christopher Mark et al**
Erl Wood Manor
Windlesham Surrey GU20 6PH (GB)

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European patent convention).

Courier Press, Leamington Spa, England.

EP 0 176 341 B1

Description

The present invention relates to an improved, shortened derivative of the tryptophan operon. The tryptophan (*trp*) operon of *Escherichia coli* has been used to investigate virtually every aspect of amino acid metabolism, operon structure and gene structure and function. Miozzari *et al.*, 1978, J. Bacteriol. 133:1457—1446 constructed a plasmid containing the *trp* promoter-operator and portions of the leader sequence and *trp*E gene. A deletion removing the sequence encoding the carboxy terminal portion of the leader peptide, the entire attenuator region of the leader sequence and the sequence encoding the amino terminal portion of *trp*E led to the production of a fused gene product, LE. The derivative of the *trp* operon produced by this deletion is useful in the production of fused gene products comprising the LE peptide and a polypeptide of research or commercial value.

The present invention comprises improved derivatives of the modified *trp* operon constructed by Miozzari. These improved derivatives were made by introducing deletions in the LE-encoding region. Two well-characterized deletions, *trp*LE1 and *trp*LE2, are especially useful. Expression vectors containing either *trp*LE1 or *trp*LE2 were constructed to drive expression of fused gene products in *Escherichia coli*. When fused to small proteins such as insulin-like growth factors I and II, these altered LE proteins represent a much smaller fraction of the fused gene product than would be the case if the original LE protein was present. This is extremely advantageous as it provides for a greater recovery of the desired protein.

The invention further comprises a novel DNA sequence encoding the gene for insulin-like growth factor II and a novel restriction fragment encoding the gene for insulin-like growth factor I. Additionally, the invention comprises expression vectors and transformants comprising the aforementioned DNA sequences.

The present invention is related generally to European Patent Publication 0036776A2, published on September 30, 1981. The aforementioned application discloses plasmids employing a modified tryptophan operon that has been useful as an expression system (Miozzari *et al.*, 1978). This deletion allows maximal derepression of the tryptophan operon and high expression of polypeptide products. The publication does not, however, disclose the present modified tryptophan operons or suggest their utility as critical components of improved expression vectors and methods.

The present invention also provides for the cloning and expression of synthetic insulin-like growth factors I and II (IGFI and IGFII) gene sequences in *Escherichia coli* and related organisms. Somatomedins, such as IGFI and IGFII, constitute a heterogeneous group of peptides with important growth-promoting effects *in vitro* as well as *in vivo*. Such peptides are believed to mediate the growth-promoting actions of growth hormone and may be useful in the treatment of: dwarfism, osteoporosis, cartilage degeneration, heart and skeletal muscle deterioration, wounds and protein and carbohydrate dysfunction. Heretofore, IGFs could be isolated from human plasma in which they are present in extremely small quantities. Research in the area of IGF function has been extremely limited because of the general paucity of available, purified IGF. The present invention overcomes this limitation since the cloning and expression of these synthetic genes by way of recombinant DNA technology provides an efficient means of producing large quantities of these important proteins and thus represents a significant advance in the technical art.

For purposes of the present invention, as disclosed and claimed herein, the following terms are as defined below.

Recombinant DNA Cloning Vector — any autonomously replicating or integrating agent, including, but not limited to, plasmids, comprising a DNA molecule to which one or more additional DNA segments can be or have been added.

Recombinant DNA Expression Vector — any recombinant DNA cloning vector into which one or more transcriptional and translational activator sequence(s) has been incorporated.

Transcriptional Activating Sequence — any DNA sequence that directs or provides for the transcription of DNA into a mRNA transcript.

Translational Activating Sequence — any DNA sequence that provides for the translation of a mRNA transcript into a polypeptide.

Leader Sequence — a segment of an operon between the transcriptional activating sequence and the structural gene(s); in the *trp* operon, a portion of the leader sequence encodes a leader peptide.

*Trp*LE — a DNA segment encoding that portion of the tryptophan operon consisting of the promoter, the operator, the sequence encoding the amino terminal portion of the leader peptide and the sequence encoding the carboxy terminal portion of *trp*E; thus, *trp*LE derives expression of a fused gene product called LE.

Functional Polypeptide — a recoverable bioactive heterologous polypeptide or precursor, a recoverable bioactive polypeptide comprising a heterologous polypeptide and a portion or whole of a homologous polypeptide, or a recoverable bioinactive fusion polypeptide comprising a heterologous polypeptide and a bioinactivating polypeptide which can be specifically cleaved.

Restriction Fragment — any linear DNA sequence generated by the action of one or more restriction enzymes.

Fused Gene Product — a recoverable heterologous polypeptide which is fused with a portion or whole of a homologous polypeptide.

Replicon — any DNA sequence that controls the replication of recombinant DNA cloning and expression vectors.

Runaway Replicon — a replicon which lacks or can be induced to lose copy number control, such loss resulting in the uncontrolled replication and an extreme increase in the copy number of DNA into which such replicon has been incorporated.

Transformation — the introduction of DNA into a recipient host cell that changes the genotype and results in a change in the recipient cell.

Transformant — a recipient host cell that has undergone transformation.

Ap^R — the ampicillin-resistant phenotype.

Km^R — the kanamycin-resistant phenotype.

EK-BGH — the enterokinase-bovine growth hormone sequence.

Detailed Description of the Invention

The present invention provides an improved, shortened derivative of the *E. coli* tryptophan operon promoter-operator/leader/tryptophan E gene sequence from which sequences encoding the carboxy terminus of the leader peptide, the attenuator region of the leader sequence and the proximal region of the *trp* E gene have been deleted, which comprises a leader/tryptophan E gene sequence contained within about 428 to about 548 deoxyribonucleotide pairs. The invention further comprises expression vectors and transformants comprising the aforementioned DNA.

As discussed above, the present invention can be used to drive expression of various functional polypeptides in *Escherichia coli*. A number of researchers have reported the use of the tryptophan operon regulatory region to express foreign genes in *E. coli*. The majority of expression vectors containing the tryptophan operon which are in use at the present time contain the *trp* leader sequence (Edman *et al.*, 1981, Nature 291:503; Miozzari *et al.*, 1978, J. Bacteriology 133:1457; Kleid *et al.*, European Patent Publication 0036 776 A2; and Goeddel *et al.*, 1980, Nature 287:411). The vectors are quite useful in expression systems but are limited in their ability to efficiently produce desired proteins due to the large percentage of *trp* leader-derived polypeptide in the fused gene product.

The *trp*LE1 and *trp*LE2 sequences of the present invention are, respectively, about 428 and about 548 base pairs. The nucleotide sequence of the *trp*LE1 coding strand is:

```
 GA ATT CAC GCT GTG GTG TTA TGG TCG GTG GTC GCT AGG
GTG CCG ACG CGC ATC TCG ACT GCA CGG TGC ACC AAT GCT
TCT GGC GTC AGG CAG CCA ATC GGA AGC TGT GGT ATG GCT
GTG CAG GTC GTA TAA TCA CCG CAT AAT TCG AGT CGC TCA
AGG CGC ACT CCC GTT CCG GAT AAT GTT TTT TGC TCC GAC
ATC ATA ACG GTT CCG GCA AAT ATT CTG AAA TGA GCT GTT
GAC AAT TAA TCA TCG AAC TAG TTA ACT AGT ACG CAA GTT
CAC GTA AAA AGG GTA TCG ACA ATG AAA GCA ATT TTC GTA
CTG AAA GGT TCA CTG GAC AGA GAT CAT TCT GTT CCG CAG
TCG GAA GCC GAC GAA ACC CGT AAC AAA GCC CGC GCT GTA
CTG CGC GCT ATT GCC ACC GCG CAT CAT GCA CAG GAA TTC
```

and the respective coding strand for *trp*LE2 is:

```
 GA ATT CAC GCT GTG GTG TTA TGG TCG GTG GTC GCT AGG
GTG CCG ACG CGC ATC TCG ACT GCA CGG TGC ACC AAT GCT
TCT GGC GTC AGG CAG CCA ATC GGA AGC TGT GGT ATG GCT
GTG CAG GTC GTA TAA TCA CCG CAT AAT TCG AGT CGC TCA
AGG CGC ACT CCC GTT CCG GAT AAT GTT TTT TGC TCC GAC
ATC ATA ACG GTT CCG GCA AAT ATT CTG AAA TGA GCT GTT
GAC AAT TAA TCA TCG AAC TAG TTA ACT AGT ACG CAA GTT
```

CAC GTA AAA AGG GTA TCG ACA ATG AAA GCA ATT TTC GTA
CTG AAA GGT TCA CTG GAC GGC AGC TAC GGC GGC GCG GTA
GGT TAT TTC ACC GCG CAT GGC GAT CTC GAC ACC TGC ATT
GTG ATC CGC TCG GCG CTG GTG GAA AAC GGT ATC GCC ACC
GTG CAA GCG GGT GCT GGT GTA GTC CTT GAT TCT GTT CCG
CAG TCG GAA GCC GAC GAA ACC CGT AAC AAA GCC CGC GCT
GTA CTG CGC GCT ATT GCC ACC GCG CAT CAT GCA CAG GAA
TTC.

wherein A is deoxyadenyl, G is deoxyguanyl, C is deoxycytidyl and T is thymidyl.

The above sequences were constructed by internally deleting various regions within the LE-encoding sequence of *trp*LE. These deletions were performed such that the translational reading phase was not altered upon subsequent ligation to a gene encoding a protein of interest.

By deleting internal regions of the LE sequence, the *Eco*RI restriction sites which bound the present *trp*LE1 and *trp*LE2 sequences were left intact for use in subsequent ligations. The aforedescribed *trp*LE1 and *trp*LE2 sequences can be conveniently modified to facilitate ligation if the *Eco*RI sites are not desired. In addition to the present LE1 and LE2 sequences, many other internal deletions which retain the desired reading phase can be constructed resulting in *trp*LE sequences that are also within the scope of the present invention. These sequences can be modified by adding, eliminating, or substituting nucleotides to alter characteristics and to provide a variety of unique or additional restriction sites. Those skilled in the art understand nucleotide chemistry and the genetic code and thus which nucleotides are interchangeable and which DNA modifications are desirable for a specific purpose.

The deletions described herein result in significantly shortened LE polypeptide products, thereby diminishing the overall energy and amino acid precursor requirements of the host and enhancing cellular economy and efficiency. Furthermore, the difference in the molecular weight of LE1 and LE2 proteins allows one the option to produce fused gene products of differing size. This is very beneficial when the expression of small polypeptides is desired. As is known in the art, small polypeptides are extremely difficult to obtain using direct expression methods; therefore, expression systems utilizing a fusion protein are often preferred to enhance recovery of the desired protein. The availability of both the LE1 and LE2 sequences, in conjunction with their adaptability, provides an opportunity to achieve greater expression results. In the case of very small peptides, a somewhat intermediately sized LE-encoding sequence, such as that in *trp*LE2, is preferred over the extremely shortened LE-encoding sequence in *trp*LE1. For example, the expression of the B chain of insulin, a very small polypeptide of 30 amino acids, was greater when fused with *trp*LE2 than with *trp*LE1. Also, in certain cases the ultimate purification of desired protein products is facilitated by reduction of the fused gene product's size and complexity.

Another attribute of the present *trp*LE sequences lies in the resultant changes in the chemical composition and amino acid sequence of the fused gene product. For example, the absence of any cysteine codons in the *trp*LE1 sequence insures that no internal disulfide bonds will be produced within the LE1 peptide. Additionally, the presence of only one methionine codon in *trp*LE1 and two methionine codons in *trp*LE2 confers great purification advantages, as fewer protein fragments are generated upon cyanogen bromide cleavage of the fused gene product and fewer proteins need to be separated than when using unimproved *trp*LE. In contrast, the known *trp*LE sequence has seven methionine codons and thus purification after cyanogen bromide cleavage of the fused gene product is difficult and time-consuming.

Expression vectors were constructed containing either the *trp*LE1 or *trp*LE2 nucleotide sequences. The sequences were individually fused to genes encoding small proteins, such as bovine growth hormone, pro-insulin, the A and B chains of insulin and IGFI and IGFII, for expression in *E. coli*. Thus, the invention further comprises recombinant DNA expression vectors for the production in *Escherichia coli* and related hosts of a heterologous polypeptide product, said vector comprising:

a) a tryptophan operon promoter-operator and shortened leader-encoding sequence;

b) a nucleotide sequence encoding a structural gene for a heterologous polypeptide, said nucleotide sequence located downstream and in translational reading phase with the tryptophan-promoter-operator and shortened leader-encoding sequence; and

c) a replicon and selectable marker that provide for vector replication and allow selection in a micro-organism transformed by the vector.

The present vectors are extremely advantageous when used to express desired proteins. The fusion of the present *trp*LE1 or *trp*LE2 sequence to a gene encoding a protein of interest allows one to produce a larger percent of the desired protein. This result is evident whenever a finite amount of product can be accumulated per bacterial cell. It follows that a decreased amount of undesirable product, that is, the LE protein, results in an increased amount of the desired protein. Thus, the present vectors, utilizing the *trp*LE1 and *trp*LE2 sequences of the present invention, are extremely advantageous, especially as applied to the

production of commercially desirable proteins.

Expression vector pCZ20 is approximately 11 kb and contains genetic sequences encoding *trp*LE1, a synthetic gene for IGFI and a runaway replicon. Plasmid pLEBGH2 is approximately 6 kb and contains genetic sequences encoding *trp*LE2 and enterokinase-linked bovine growth hormone (EK-BGH). Plasmids pCZ20 and pLEBGH2 can be isolated from *E. coli* K12 RV308/pCZ20 and *E. coli* K12 RR1/pLEBGH2 respectively. The strains have been deposited and made part of the stock culture collection of the Northern Regional Research Laboratory, Peoria, Illinois 61604 and are available to the public as a source and stock reservoir of the plasmids under the respective accession numbers NRRL B-15881 and B-15882. Detailed restriction site and function maps of plasmids pCZ20 and pLEBGH2 are respectively presented in Figures 1 and 2 of the accompanying drawings. Figure legends appear in the text. For purposes of the present application, Fig. 1 and all subsequent figures are not necessarily drawn to scale.

For convenience and ease of construction, an ~0.43 kb *Eco*RI *trp*LE1-containing restriction fragment can be isolated from plasmid pCZ20 and used as a starting material to construct numerous derivative plasmids. For example, the construction of plasmid pLEBGH1 is performed by first digesting plasmid pLEBGH2 with *Eco*RI restriction enzyme and then separating the ~5.5 kb *Eco*RI fragment (containing the gene encoding the EK-BGH polypeptide and an *E. coli* replicon) from the ~0.55 kb *Eco*RI restriction fragment containing the *trp*LE2 sequence. Subsequently, the ~0.43 kb *Eco*RI restriction fragment containing the *trp*LE1 sequence is ligated to the ~5.5 kb *Eco*RI restriction fragment to form plasmid pLEBGH1. Additionally, *trp*LE1 and *trp*LE2 can independently be ligated to an *Eco*RI restriction fragment containing an *E. coli* replicon, a functional antibiotic resistance gene and a synthetic IGFII-encoding gene to form plasmids pIGF201 and pIGF202, respectively.

Plasmids pIGF201 and pIGF202 were constructed to express the IGFII polypeptide under the control of *trp*LE1 and *trp*LE2, respectively. Since the nucleotide sequence of the gene encoding IGFII had not, prior to this invention, been isolated or determined, a synthetic gene was constructed. The synthesis of the coding region of the IGFII gene was completed given the knowledge that IGFII consists of 67 amino acids in the following sequence:

```
Ala Tyr Arg Pro Ser Glu Thr Leu Cys Gly Gly Glu Leu
Val Asp Thr Leu Gln Phe Val Cys Gly Asp Arg Gly Phe
Tyr Phe Ser Arg Pro Ala Ser Arg Val Ser Arg Arg Ser
Arg Gly Ile Val Glu Glu Cys Cys Phe Arg Ser Cys Asp
Leu Ala Leu Leu Glu Thr Tyr Cys Ala Thr Pro Ala Lys
Ser Glu
```

(see, for example, Blundell *et al.*, 1980, Nature 287:781—787).

From the above amino acid sequence, a corresponding synthetic gene sequence has been invented, subject to a number of specific non-obvious criteria, and oligonucleotide fragments synthesized which, when assembled, form a synthetic gene coding for IGFII. The fragments have been hybridized and ligated in predetermined stages to construct the IGFII gene in two portions. These two portions have been cloned into plasmid pBR322 so as to produce a full length IGFII gene flanked only by plasmid pBR322 DNA. The *trp*LE1 and *trp*LE2-containing fragments were inserted into the IGFII-containing pBR322 vector to maximize expression of the gene in *E. coli* and related organisms. A fused gene product comprising IGFII has been expressed in *E. coli*.

Given the above amino acid sequence and the degenerative nature of the genetic code, it is possible to predict numerous nucleotide sequences which would code for IGFII. In the inventive determination of an optimum sequence from the large number of sequences possible, several non-obvious criteria have been applied. Firstly, the trinucleotide codons used in the sequence were those known to be accepted or preferred by *E. coli*. Secondly, different restriction enzyme recognition sites were placed at the termini of the molecule so as to allow insertion into a plasmid in a desired orientation. Moreover, it was decided to select sites which allowed the use of well-understood cloning vectors, such as plasmid pBR322. In fact, *Eco*RI and *Bam*HI recognition sites were selected and introduced at the 5′ and 3′ ends, respectively, of the IGFII-encoding sequence. Thirdly, a series of restriction endonuclease recognition sites were strategically placed along the molecule to enable the gene to be specifically dissected to aid characterization and, possibly, mutagenesis. In particular, a *Pst*I site was introduced at a central location in the gene. Also, the placement of restriction enzyme recognition sites allowed the two portions of the molecule to be cloned in stages. Fourthly, since the protein ultimately expressed in bacterial cells was to be in the form of a fusion product, it was desirable to have a means of cleaving the IGFII portion from such a fusion product. The codon specifying the amino acid methionine was introduced near the end of the gene corresponding to the IGFII amino-terminus in order for there to be a methionine present in the fusion product to serve as a substrate for cyanogen bromide cleavage. Fifthly, two stop codons were introduced at the end of the gene to prevent read-through translation.

The particular preferred sequence selected for the coding region of the synthetic IGFII gene is as follows:

```
GCTTATCGAC CGTCTGAAAC TCTGTGCGGC GGCGAACTGG
TTGACACTCT GCAGTTCGTT TGCGGCGACC GTGGCTTCTA
CTTCTCTCGT CCGGCTTCTC GTGTTTCTAG ACGTTCTCGT
GGCATCGTTG AAGAATGCTG CTTCCGCTCT TGCGACCTGG
CTCTGCTGGA AACTTACTGC GCTACTCCTG CTAAATCTGA A
```

wherein A is deoxyadenyl, G is deoxyguanyl, C is deoxycytidyl and T is thymidyl.

The present vectors for expressing functional polypeptides in *E. coli* represent a significant technical advance. The aforedescribed *trp*LE1 and *trp*LE2 DNA sequences can be used for the universal expression in *E. coli* and related organisms of any polypeptide-encoding gene. While specific embodiments of the invention are shown and described in the following examples, many variations are possible. For example, the present invention is in no way limited to the use of a particular polypeptide-encoding gene since the choice of a specific sequence is not critical to the operability of the present invention. Genes coding for a functional polypeptide can be substituted for the EK-BGH-, IGFI, IGFII- insulin A chain-, insulin B chain- and proinsulin-encoding genes presently exemplified. Such coding sequences include, but are not limited to, sequences that code for human growth hormone, human pregrowth hormone, porcine growth hormone, mammalian growth hormone, avian growth hormone, growth hormone releasing factor, human pre-proinsulin, human and non-human interferon, viral antigen, urokinase, tissue plasminogen activator, interleukin II, any peptide hormone, any enzyme or virtually any other polypeptide with research or commercial value.

The present vectors are not limited to the use of a specific replicon from a particular *E. coli* plasmid. Although the *E. coli* replicon exemplified in the majority of the present vectors is from plasmid pBR322, other *E. coli* replicon-containing fragments can be obtained from, for example, plasmid pBR324 (disclosed in Bolivar, F., 1978, Gene 4:121), pBR324 (disclosed in Soberon, X., 1980, Gene 9:287), plasmid pKN402 (disclosed in Uhlin *et al.*, 1979, Gene, 6:91—106), or the like to produce novel vectors. Those skilled in the art will understand that ligation of these, or any other, *E. coli* replicon-containing fragments results in vectors that are within the scope of the present invention.

The expression vectors of this invention can be applied to a wide range of host organisms, for example, Gram-negative prokaryotic organisms such as *Escherichia coli*, *E. coli* K12, *E. coli* K12 RV308, *E. coli* K12 HB101, *E. coli* K12 C600R_k-M_k-, *E. coli* K12 RR1, *E. coli* K12 MM294 and the like. While all of the embodiments of the present invention are useful, some of the vectors and transformants are preferred. Preferred vectors are pCZ20, pLEBGH1, pLEBGH2, pIGF201, pIGF202, pIALE1, pIALE2, pIBLE1, pIBLE2, pPILE1 and pPILE2. Preferred transformants include *E. coli* K12 RV308/PCZ20, *E. coli* K12 RV308/pLEBGH2, *E. coli* K12 RV308/pLEBGH1, *E. coli* K12 RV308/pIGF201, *E. coli* K12 RV308/pIGF202, *E. coli* K12 MM294/pIGF201, *E. coli* K12 MM294/pIGF202, *E. coli* K12 RV308/pIALE1, *E. coli* K12 RV308/pIBLE2, and *E. coli* K12 RV308/pPILE1. Moreover, of this preferred group, plasmids pCZ20, pLEBGH1, pIGF201 and pIBLE2 and transformants *E. coli* K12 RV308/PCZ20, *E. coli* K12 RV308/pLEBGH1, *E. coli* K12 RV308/pIGF201 and *E. coli* K12 RV308/PIBLE2 are especially preferred.

The following examples further illustrate the invention disclosed herein. Both an explanation of and the actual procedures for constructing the invention are described where appropriate.

Brief Description of the Figures

Figure 1 is a restriction site and function map of plasmid pCZ20.
Figure 2 is a restriction site and function map of plasmid pLEBGH2.
Figure 3 is a restriction site and function map of plasmid pIGF201.
Figure 4 is a restriction site and function map of plasmid pIGF202.
Figure 5 is a restriction site and function map of plasmid pIALE1.
Figure 6 is a restriction site and function map of plasmid pIBLE2.
Figure 7 is a restriction site and function map of plasmid pPILE1.
Figure 8 is the result of polyacrylamide gel electrophoresis of cell protein containing the IGFII fusion product.

Example 1

Culture of *Escherichia coli* RV308/pCZ20 and Isolation of Plasmid pCZ20

A. Culture of *Escherichia coli* RV308/pCZ20

One hundred ml of TY broth (10 g tryptone, 5 g yeast extract and 5 g NaCl per liter) containing 50 µg/ml kanamycin sulfate were inoculated with a culture of *E. coli* RV308/pCZ20 (NRRL B-15881) and incubated

with shaking for ~16 hours at 20—25°C. The 100 ml of culture were then transferred to a flask containing 900 ml TY broth containing 50 µg/ml kanamycin. The diluted culture was then incubated with shaking at 37°C for 2—3 hours. The 37°C temperature of incubation induced high plasmid copy number.

B. Isolation of Plasmid pCZ20

The cells were pelleted by centrifugation (4°C at 10,000 rpm for 5 minutes) and the pellet resuspended in 20 ml of a solution containing 25 mM Tris-HCl, pH 8; 10 mM EDTA; and 50 mM glucose and supplemented with 2 mg/ml lysozyme. The resuspended cells were incubated on ice for 15 minutes, and then 40 ml of a solution containing 1% SDS and 0.2N NaOH were added and mixed. After the cells were completely lysed, 30 ml of cold 3M NaOAc, pH 4.8, were added, mixed and the resulting solution incubated on ice for one hour. This solution was then centrifuged at 20,000 rpm for 30 minutes. After centrifugation, the pellet was discarded and 3 volumes cold, absolute ethanol were added to the supernatant. The resulting mixture was chilled at −70°C for 10—20 minutes and then centrifuged at 10,000 rpm for 10 minutes to pellet the DNA.

The DNA pellet was resuspended in 10 ml of TE buffer (10 mM Tris-HCl, pH 7.5 and 1 mM EDTA), and then 0.1 ml of a 5 mg/ml RNAse A solution and 10 µl of a 2500 U/ml RNAse T solution were added. After the RNAses were mixed in with the solution, the solution was incubated at 65°C for 20 minutes. Next, 30 g of CsCl were added and the volume adjusted to 38 ml, total, with TE buffer. Two ml of ethidium bromide were added, and ultracentrifugation at 49,000 rpm for 17 hours in a vertical rotor was performed to band the plasmid DNA.

After the plasmid band was removed from the centrifugation tube, the ethidium bromide was extracted with isopropanol (saturated with CsCl and H_2O) and the CsCl removed by dialysis against TE buffer. The resultant plasmid pCZ20 DNA was suspended in TE buffer at a concentration of 1 µg/ml and stored at −20°C. A restriction site and function map of plasmid pCZ20 is presented in Fig. 1 of the accompanying drawings.

Example 2

Isolation of the ~0.43 kb *Eco*RI Restriction Fragment Encoding *trp*LE1 From Plasmid pCZ20

About 30 µl (30 µg) of the plasmid pCZ20 DNA, isolated above, were added to 10 µl 10X *Eco*RI buffer (1.5 M Tris-HCl, pH 7.2; 500 mM NaCl; and 10 mM dithiothreitol), 2 µl *Eco*RI restriction enzyme (~60 units) and 58 µl H_2O. After mixing, the reaction was placed in a 37°C water bath for one hour, and then the solution was electrophoresed on a 1% agarose gel until the desired ~0.43 kb *Eco*RI fragment was clearly separated from the other digestion products. Visualization of the electrophoresed DNA was accomplished by staining the gel in a dilute solution (0.5 µg/ml) of ethidium bromide and exposing the stained gel to long-wave UV light. After locating the desired fragment, a small slit was made in the gel and a small piece of Schleicher and Schuell (Keene, NH 03431) NA-45 DEAE membrane was placed in the slit. Upon further electrophoresis the DNA non-covalently bound to the DEAE membrane. After the desired fragment was bound to the DEAE membrane, the membrane was removed and rinsed with low salt buffer (150 mM NaCl; 0.1 mM EDTA; and 20 mM Tris-HCl, pH 8). Next, the membrane was placed in a small tube and immersed in high salt buffer (1M NaCl; 0.1 mM EDTA; and 20 mM Tris-HCl, pH 8) and then incubated at 65°C for one hour to remove the DNA from the DEAE paper. After incubation, the incubation buffer was collected and the membrane rinsed with high salt buffer. The rinse solution was pooled with the incubation buffer before collecting the desired DNA fragment.

The volume of the high salt-DNA solution was adjusted so that the NaCl concentration was 0.25M, and then three volumes of cold, absolute ethanol were added. The resulting solution was mixed and placed at −70°C for 10—20 minutes. After chilling, the solution was centrifuged at 15,000 rpm for 15 minutes. After another precipitation to remove residual salt, the DNA pellet was rinsed with ethanol, dried, resuspended in 20 µl of TE buffer and constituted ~0.25 µg of the desired *trp*LE1-encoding *Eco*RI restriction fragment.

Example 3

Culture of *Escherichia coli* RV308/pLEBGH2 and Isolation of Plasmid pLEBGH2

A. Culture of *Escherichia coli* RV308/pLEBHG2

One hundred ml of TY containing 50 µg/ml ampicillin and 100 µg/ml tryptophan were inoculated with a culture of *E. coli* RV308/pLEBGH2 (NRRL B-15882) and then incubated at 37°C for 16 hours with shaking. The culture was then diluted to 1 liter with TY broth containing 50 µg/ml ampicillin and 100 µg/ml tryptophan and incubation continued until the optical density at 550 nanometers was ~0.5 absorbance units. When this density was reached, one gram of uridine was added to the flask. The uridine was incubated with the cells for 15 minutes, and then 170 mg of chloramphenicol were added. Incubation of the culture at 37°C with shaking was continued for ~16 hours. The uridine and chloramphenicol additions are a well-known means of amplifying certain *E. coli* plasmids.

B. Isolation of Plasmid pLEBGH2

Plasmid pLEBGH2 was isolated from the above culture in substantial accordance with the procedure of Example 1B. The resultant plasmid, pLEBGH2, was suspended in TE buffer at a concentration of 1 µg/µl and stored at −20°C. A restriction site and function map of plasmid pLEBGH2 is presented in Fig. 2 of the accompanying drawings.

Example 4

*Eco*RI Digestion of Plasmid pLEBGH2

About 30 µl of plasmid pLEBGH2 DNA (from Example 3) were digested with *Eco*RI and the reaction products isolated in substantial accordance with the procedures of Example 2. Two *Eco*RI fragments were isolated and purified: 1) the ~0.55 kb *Eco*RI fragment encoding the *trp*LE2 sequence; and 2) the ~5.5 kb *Eco*RI fragment encoding EK-BGH, the pBR322 replicon and the β-lactamase gene. About 0.65 µg of the *trp*LE2-encoding fragment was obtained and suspended in 20 µl of TE, and about 6.5 µg of the ~5.5 kb fragment were obtained and suspended in 100 µl of phosphatase buffer (10 mM Tris-HCl, pH 8; 1 mM $MgCl_2$; and 0.01 mM $ZnCl_2$).

Example 5

Construction of *Escherichia coli* RV308/pLEBGH1

A. Construction of Plasmid pLEBGH1

The ~5.5 kb *Eco*RI restriction fragment suspended in phosphatase buffer (from Example 4) was incubated at 65°C for 5 minutes, and then 1 µl (7 Units Boehringer-Mannheim) of calf-intestinal alkaline phosphatase was added, mixed and incubation at 65°C continued for 5 more minutes. The incubation was followed by a 30 minute incubation at 60°C, after which the reaction mix was extracted once with phenol:$CHCl_3$ (50:50) and once with $CHCl_3$. After the extractions, the reaction mix was made 0.3M in NaOAc; three volumes of ethanol were added; and, after mixing and chilling the solution to −70°C, the solution was centrifuged to pellet the phosphatased fragment. The DNA pellet obtained was suspended in 25 µl of TE buffer and constituted ~5 µg of the phosphatased ~5.5 kb *Eco*RI restriction fragment of plasmid pLEBGH2.

About 1 µl of the phosphatased ~5.5 kb *Eco*RI fragment was mixed with 4 µl of the purified ~0.43 kb *Eco*RI fragment encoding *trp*LE1 obtained in Example 2, 2 µl of 5 mM ATP, 2 µl of 10 mM dithiothreitol, 2 µl of 10X ligase buffer (660 mM Tris-HCl, pH 8 and 60 mM $MgCl_2$), 8 µl of water and 1 µl (800 Units New England Biolabs) of T4 DNA ligase. The reaction was incubated at 20°C for two hours. The ligated DNA was subsequently transformed into *E. coli* K12 RV308.

B. Construction of *Escherichia coli* RV308/pLEBGH1

1. Preparation of Frozen, Competent *Escherichia coli* K12 RV308 and *Escherichia coli* K12 MM294

Five ml portions of TY broth were individually inoculated with *E. coli* K12 RV308 (NRRL B-15624) and *E. coli* K12 MM294 (NRRL B-15625) and the resulting cultures incubated at 37°C overnight with shaking. The overnight cultures were diluted with TY broth to a final volume of 1 liter, resulting in an optical density reading (600 nanometers) of each culture that was ~0.1 absorbance units. The incubation at 37°C with shaking was continued until the optical density reading (600 nm) reached the 0.55—0.65 absorbance units range, and then the cells were collected by centrifugation.

The cell pellets were individually resuspended in 500 ml of chilled 50 mM $CaCl_2$ and the resulting mixtures incubated on ice for 15—30 minutes. The cells were then collected by centrifugation, and the resulting pellets were resuspended in 20 ml of a cold solution of 20% glycerol in 50 mM $CaCl_2$. The cell mixtures were then aliquoted in 0.2 ml portions into pre-chilled tubes, which were immediately placed and stored at −70°C. The cells prepared by this procedure remain viable and competent for transformation for up to one year.

2. Transformation

One of the tubes containing the competent *E. coli* K12 RV308 cells was removed from storage at −70°C, thawed and mixed with the ligated DNA of Example 5A. The cell-DNA mix was incubated on ice for one hour. The cells were then collected, the supernatant discarded and the pellet resuspended in 0.5 ml of TY broth supplemented with tryptophan at 100 µg/ml. After incubation for 30 minutes at 37°C, the cells were plated on TY plates supplemented with 50 µg/ml ampicillin and 100 µg/ml tryptophan. The plates were incubated at 37°C overnight.

3. Analysis

Since the ligation products prepared in Example 5A comprised both the desired plasmid pLEBGH1 and other undesired plasmids, the individual transformants were analyzed in order to determine in which transformants plasmid LEBGH1 was present.

Since the entire DNA sequence of plasmid pLEBGH1 could be predicted, the isolated plasmid DNA from the transformed *E. coli* cells was cleaved with different restriction enzymes to determine, by electrophoresis and gel analysis, whether the reaction products were those predicted for plasmid pLEBGH1. The proteins expressed by the transformants were analyzed in the absence of tryptophan and the desired transformants identified by their production of LE1-EK-BGH (38,000 daltons). LE1-EK-BGH and LE2-EK-BGH also comprise protein encoded by pBR322 DNA. In this manner, *E. coli* K12 RV308/pLEBGH1 transformants were identified and isolated.

Example 6

Construction of IGFII-Encoding DNA

The synthesis of the coding region of the IGFII gene was accomplished by the following generalized procedure: A) 38 single-stranded deoxyribooligonucleotides, each containing between 9 and 15 deoxyribonucleotides, were synthesized by the improved phosphotriester method; B) some of the 38 single-stranded DNA molecules were phosphorylated; and C) a series of annealing and ligating reactions were done to form two double-stranded DNA molecules, each comprising about half of the coding region of the gene.

The two fragments were formed above were ultimately inserted into plasmid pBR322 to construct the entire IGFII coding sequence on a single DNA molecule (see Example 7). A more detailed description of steps A—C is now provided.

A. Synthesis of Single-Stranded DNA Fragments

The 38 deoxyribooligonucleotides listed below in Table 1 were synthesized by the improved phosphotriester method of Hsiung *et al.*, 1983, Nucleic Acids Research, 11:3227. A variety of DNA-synthesizing instruments are also available, well-known and suitable for synthesizing the single-stranded fragments.

Table 1

#	Sequence	Size
1	AATTCATGGCT	11mer
2	TATCGACCGTCT	12mer
3	GGCGGCGAACTG	12mer
4	AAGCCACGGT	10mer
5	GTTGACACTCTG	12mer
6	CGCCGCAAACGA	12mer
7	GCTTCTACTTC	11mer
8	TCTCGTCCGG	10mer
9	CTTCTCGTGTTT	12mer
10	CTAGACGTTC	10mer
11	TCGTGGCAT	9mer
12	CGTTGAAGAATG	12mer
13	TCTTGCGACCTG	12mer
14	CAGTTCGTTTGC	12mer
15	GCTCTGCTGG	10mer
16	AAACTTACTGC	11mer
17	GCTACTCCTGCT	12mer
18	AAATCTGAATAATAG	15mer
19	CGATAAGCCATG	12mer
20	GTTTCAGACGGT	12mer
21	CAACCAGTTCGC	12mer
22	ACTGCAGAGTGT	12mer
23	CTGCTTCCGC	10mer
24	GGCGACCGTG	10mer
25	GAGAGAAGTAG	11mer
26	GAAGCCGGAC	10mer
27	CTAGAAACACGA	12mer
28	ACGAGAACGT	10mer

Table 1 continued

#	Sequence	Size
29	AACGATGCC	9mer
30	AAGCAGCATTCTTC	14mer
31	TCGCAAGAGCGG	12mer
32	CAGAGCCAGG	10mer
33	AAGTTTCCAG	10mer
34	AGTAGCGCAGT	11mer
35	AGATTTAGCAGG	12mer
36	GATCCTATTATTC	13mer
37	GAAACTCTGTGC	12mer
38	CGCCGCACAGA	11mer

B. Phosphorylation

After purifying each oligonucleotide by thin layer chromatography and reversed phase high pressure liquid chromatography, certain of the 38 single-stranded DNA fragments of Table 1 were phosphorylated according to the teaching of Hsiung *et al.*, 1983, in order to facilitate the ligation and construction of the IGFII gene-encoding DNA fragments. Some of the fragments are depicted as having "^{32}P" at their 5′ end, since tracer amounts of [γ-^{32}P]-ATP were used in the phosphorylation reactions.

C. Annealing and Ligation

Fragments 1, 2, 3, 19, 20, 21, 37 and 38 were annealed and ligated to form Duplex I:

```
5'  HO-AATTCATGGCTTATCGACCGTCTGAAACTCTGTGCGGCGGCGAACTG-OH      3'
        ||||||||||||||||||||||||||||||||||||||||||||||||
3'      HO-GTACCGAATAGCTGGCAGACTTTGAGACACGCCGCCGCTTGACCAAC-32P  5'
```

Fragments 5, 14, 24, 22, 6 and 4 were annealed and ligated to form Duplex II:

```
5'  32P-GTTGACACTCTGCAGTTCGTTTGCGGCGACCGTG-OH      3'
        ||||||||||||||||||||||||||||||||
3'  HO-   TGTGAGACGTCAAGCAAACGCCGCTGGCACCGAA-32P 5'
```

Fragments 7, 8, 9, 25, 26 and 27 were annealed and ligated to form Duplex III:

```
5' 32P-GCTTCTACTTCTCTCGTCCGGCTTCTCGTGTTT-OH     3'
       ||||||||||||||||||||||||||||||||
3'      HO-GATGAAGAGAGCAGGCCGAAGAGCACAAAGATC-OH 5'
```

DNA duplex molecules I, II and III were then mixed and treated with T4 DNA ligase to form a double-stranded molecule with an *Eco*RI overlap at one end of the molecule and a *Xba*I overlap at the other end. The ligation product of the reaction involving duplexes I, II and III was purified on a 10% polyacrylamide gel and constitutes about half of the IGFII coding region.

The remainder of the IGFII coding region was synthesized in like manner. Thus, fragments 10, 11, 12, 28, 29 and 30 were annealed and ligated to form Duplex IV:

```
5' HO-CTAGACGTTCTCGTGGCATCGTTGAAGAATG-OH        3'
       ||||||||||||||||||||||||||||||
3'     HO-TGCAAGAGCACCGTAGCAACTTCTTACGACGAA-32P 5'
```

Fragments 23, 13, 15, 16, 31, 32, 33 and 34 were annealed and ligated to form Duplex V:

```
5' 32P-CTGCTTCCGCTCTTGCGACCTGGCTCTGCTGGAAACTTACTGC-OH        3'
                  ||||||||||||||||||||||||||||||||||||
3'        OH-GGCGAGAACGCTGGACCGAGACGACCTTTGAATGACGCGATGA-32P 5'
```

Fragments 17, 18, 35 and 36 were annealed and ligated to form Duplex VI:

```
5' 32P-GCTACTCCTGCTAAATCTGAATAATAG-OH     3'
                  ||||||||||||||||||||
3'        HO-GGACGATTTAGACTTATTATCCTAG-OH 5'
```

Duplex DNA molecules IV, V and VI were then ligated together to form the remaining portion of the IGFII coding region. The ligation produced a DNA molecule with an *Xba*I overlap at one end and a *Bam*HI overlap at the other. The ligation product was purified on a 10% polyacrylamide gel.

Example 7

Construction of Plasmids pIGF201 and pIGF202

A. Construction of Plasmid pIGF2

Five µg of plasmid pBR322 are dissolved in 5 µl of TE buffer, and 2 µl 10X *Bam*Hi buffer (1.5M NaCl; 60 mM Tris-HCl, pH 7.9; 60 mM $MgCl_2$; and 1 mg/ml BSA), 1 µl *Bam*HI restriction enzyme (~10 Units) and 12 µl H_2O are added, gently mixed and incubated at 37°C for 2 hours. After the incubation, the *Bam*HI-digested DNA is precipitated and then resuspended in 2 µl 10X *Eco*RI buffer, 1 µl *Eco*RI restriction enzyme (~10 Units) and 17 µl H_2O. After gentle mixing, the reaction is incubated at 37°C for 2 hours.

The *Eco*RI- and *Bam*HI-digested plasmid pBR322 DNA is extracted once with phenol-$CHCl_3$ (50:50), followed by extraction with $CHCl_3$ alone. The DNA is precipitated by making the mixture 0.3M in NaOAc, adding 2.5—3 volumes ethanol, mixing, chilling to −70°C and centrifuging. The DNA pellet constitutes ~5 µg of the *Eco*RI- and *Bam*HI-digested plasmid pBR322 DNA. After the DNA is suspended in 25 µl of TE buffer, it is stored at −20°C for subsequent ligation to the synthetic IGFII-encoding gene fragments prepared in Example 6.

One µl of the *Eco*RI- and *Bam*HI-digested plasmid pBR322 is added to 0.6 picomoles each of the *Eco*RI-*Xba*I and *Xba*I-*Bam*HI IGFII-encoding fragments generated in Example 6. The DNA molecules are ligated in substantial accordance with the ligation procedure of Example 5. The ligated DNA was used to transform *E. coli* RV308 in substantial accordance with the transformation procedure taught in Example 5. The desired transformants were identified by their predicted ampicillin-resistant, tetracycline-sensitive phenotype and by analysis of their plasmid DNA. The transformants thus identified constituted the desired *E. coli* RV308/pIGF2. Plasmid pIGF2 was prepared and purified in substantial accordance with the teaching of Example 3.

B. Digestion and Dephosphorylation of Plasmid pIGF2

Five µl of plasmid pIGF2, isolated above, were digested with *Eco*RI in a total volume of 50 µl in substantial accordance with the procedure of Example 4. The *Eco*RI-digested plasmid pIGF2 obtained was pelleted by precipitation, resuspended in 100 µl of phosphatase buffer and then treated with calf-intestinal alkaline phosphatase in substantial accordance with the procedure of Example 5A. After removing the phosphatase enzyme, the dephosphorylated *Eco*RI-digested plasmid pIGF2 was suspended in 25 µl of TE buffer and stored at −20°C.

C. Ligation and Transformation

One µl of the dephosphorylated *Eco*RI-digested plasmid pIGF2 was added to 4 µl of the *trp*LE1-encoding *Eco*RI restriction fragment isolated in Example 2 and ligated in substantial accordance with the teaching of Example 5A. The ligated DNA constituted the desired plasmid pIGF201. Plasmid pIGF202 was constructed in like manner, by using 4 µl of the *trp*LE2-encoding *Eco*RI restriction fragment prepared in Example 4 in place of the *trp*LE1-encoding *Eco*RI fragment.

The ligated DNA constituting plasmids pIGF201 and pIGF202 DNA was individually used to transform both *E. coli* K12 RV308 and *E. coli* K12 MM294 in accordance with the transformation procedure of Example 5B2. The desired transformants, *E. coli* K12 RV308/pIGF201, *E. coli* K12 MM294/pIGF201, *E. coli* K12 RV308/pIGF202 and *E. coli* K12 MM294/pIGF202, were identified by analysis of their plasmid DNA and protein production in the absence of tryptophan.

Example 8

Construction of Insulin A Chain, Insulin B Chain, and Proinsulin Expression Vectors

A. Isolation of Structural Genes

1. Isolation of the Insulin A Chain Structural Gene

The insulin A chain structural gene was obtained from plasmid pIA1; the construction of which is disclosed in Goeddel *et al.*, 1979, Proc. Nat. Acad. Sci. USA, 76:106. Approximately 50 µg of plasmid pIA1

were digested with *Bam*HI and *Eco*RI in substantial accordance with the teaching of Example 7A, and the desired ~0.425 kb insulin A chain structural gene-containing *Eco*RI-*Bam*HI restriction fragment was isolated by polyacrylamide gel electrophoresis and electroelution. Approximately 0.5 µg of the desired fragment was obtained, suspended in 25 µl of TE buffer and stored at −20°C.

2. Isolation of the Insulin B Chain Structural Gene

The insulin B chain structural gene was obtained from plasmid pIB1; the construction of which is disclosed in Goeddel *et al.*, 1979. Approximately 50 µg of plasmid pIB1 were digested with *Bam*HI and *Eco*RI in substantial accordance with the teaching of Example 7A, and the desired ~0.453 kb insulin B chain structural gene-containing *Eco*RI-*Bam*HI restriction fragment was isolated by polyacrylamide gel electrophoresis and electroelution. Approximately 0.5 µg of the desired fragment was obtained, suspended in 25 µl of TE buffer and stored at −20°C.

3. Isolation of the Proinsulin Structural Gene

The proinsulin structural gene was obtained from plasmid pHI7Δ4Δ1; the construction of plasmid pHI7Δ4Δ1 is disclosed in Examples 2 and 4 of European Patent Publication No. 105608A1, published April 18, 1984. Approximately 50 µg of plasmid pHI7Δ4Δ1 were digested with *Bam*HI and *Eco*RI in substantial accordance with the teaching of Example 7A, and the desired ~0.705 kb proinsulin structural gene-containing *Eco*RI-*Bam*HI restriction fragment was isolated by polyacrylamide gel electrophoresis and electroelution. Approximately 0.5 µg of the desired fragment was obtained, suspended in 25 µl of TE buffer and stored at −20°C.

B. Ligations

The following ligations were done in substantial accordance with the procedure of Example 5A. Six different ligations were performed in order to construct both *trp*LE1 and *trp*LE2 derivatives of each of the three structural genes isolated in Example 8A. Each ligation involved: 1 µl of the *Eco*RI- and *Bam*HI-digested plasmid pBR322 DNA prepared in Example 7A; 4 µl of either the *trp*LE1- or *trp*LE2-encoding *Eco*RI restriction fragments prepared in Examples 2 and 4, respectively; and 4 µl of one of the *Eco*RI-*Bam*HI structural gene-encoding restriction fragments isolated in Example 8A. Thus, plasmids pIALE1 and pIALE2 contain the *trp*LE1- and *trp*LE2-insulin A chain derivative genes, respectively; plasmids pIBLE1 and pIBLE2 contain the *trp*LE1- and *trp*LE2-insulin B chain derivative genes, respectively; and plasmids pPILE1 and pPILE2 contain the *trp*LE1- and *trp*LE2-proinsulin derivative genes, respectively.

C. Transformations

Each of the six ligations of Example 8B were individually used to transform *E. coli* K12 RV308 in substantial accordance with the procedure of Example 5B2. The desired transformants were identified by their ampicillin-resistant, tetracycline-sensitive phenotype and by analysis of the transformant's plasmid DNA.

Example 9
Preparation and Analysis of Cell Extracts

A. Cell Culture

Overnight cultures of *E. coli* K12 RV308/pIGF201, *E. coli* K12 RV308/pIGF202, *E. coli* K12 RV308/pIGF203, *E. coli* K12 MM294/pIGF201, *E. coli* K12 MM294/pIGF202 and *E. coli* K12 MM294/pIGF203 were prepared in TY media supplemented with 100 µg/ml each of tryptophan and ampicillin. Plasmid pIGF203 has the unimproved *trp*LE sequence in front of and in reading phase with the IGFII structural gene but is identical to plasmids pIGF201 and pIGF202 in all other respects. The overnight cultures were grown to an optical density at 550 nm of ~4.5 absorbance units. The overnight cultures were diluted with M9 media* (without added tryptophan); several dilutions of each overnight culture were made in order to observe protein synthesis occurring in the cells both in early log phase and in early stationary phase. In this manner, 12 distinct cultures were prepared. The cultures were then incubated at 37°C for 4—8 hours.

B. Preparation of Cell-Extracts

The cultures were grown in the absence of tryptophan for varying lengths of time. Then, 1 ml of each culture was removed; the optical density at 550 nm was determined, and the cells were collected by centrifugation.

The cell pellet was resuspended in the amount of loading buffer (0.125M Tris-HCl, pH 6.8; 2% SDS; 3% glycerol; 0.005% bromphenol blue; and 6M Urea) necessary to form a suspension at 20 absorbance units per ml, as determined by the optical density measured before collecting the cells. The sample was boiled

*M9 media is, per liter of H_2O: 10 g of $Na_2HPO_4 \cdot 7H_2O$; 3 g of KH_2PO_4; 0.5 g of NaCl, and 1 g of NH_4Cl. The media was supplemented with: 400 µl 1M $MgSO_4$; 40 µl 1M $CaCl_2$; 12.5 ml 20% casamino acids; 12.5 ml 40% glucose; and 10 mg/ml ampicillin.

5—10 minutes, and then 7 µl of each sample were loaded onto a SDS-polyacrylamide gel (SDS-PAGE) and electrophoresed with ~50 mA current until the bromphenol blue was near the bottom of the gel.

The gel was stained overnight in Fairbank's A stain: 0.5 g Coomassie brilliant blue; 250 ml isopropanol; 100 ml glacial acetic acid; and 650 ml H_2O. The gel was then stained 3—4 hours in Fairbank's B stain: 0.05 g Coomassie brilliant blue; 100 ml isopropanol; 100 ml glacial acetic acid; and 800 ml H_2O. The gel was destained in 10% acetic acid and then photographed.

The optical density readings at 550 nm, taken before cell harvest, are shown in Table II. The picture of the stained gel is presented in Figure 8 of the accompanying drawings.

Table II

E. coli culture	Lane # (Figure 8)	O.D. 550 (absorbance)	Incubation Time (hours after dilution)
RV308/pIGF201	1	0.65	4.5
RV308/pIGF202	2	0.48	4.5
RV308/pIGF203	3	0.51	6.0
RV308/pIGF201	4	2.9	8.0
RV308/pIGF202	5	2.3	8.0
RV308/pIGF203	6	0.80	8.0
MM294/pIGF201	7	0.67	4.5
MM294/pIGF202	8	0.46	4.5
MM294/pIGF203	9	0.27	6.0
MM294/pIGF201	10	2.4	8.0
MM294/pIGF202	11	1.9	8.0
MM294/pIGF203	12	0.6	8.0

Lane 13 (Figure 8) contains Bio-Rad's (32nd and Griffin, Richmond, CA 94804) Low Molecular Weight Protein Standards.

The data presented in Table II show that cells containing plasmid pIGF203 grew more slowly than cells containing either plasmid pIGF201 or plasmid pIGF202. Thus, the presence and expression of a shortened *trp*LE sequence has a lesser growth-retarding effect on *E. coli* K12 MM294 and *E. coli* K12 RV308 than does the presence and expression of the unimproved *trp*LE. Figure 8, lanes 7—12, demonstrates that *trp*LE1 and *trp*LE2 drive greater production and accumulation of IGFII in *E. coli* K12 MM294 than does unimproved *trp*LE. Cell growth rate, product synthesis and product accumulation are all important factors in fermentation. The above evidence supports the claimed improved nature of the present invention by demonstrating that cells expressing *trp*LE1-IGFII or *trp*LE2-IGFII fusion polypeptides grow faster than cells expressing unimproved *trp*LE-IGFII fusion polypeptides. Additionally, the desired product represents a greater percentage of total cell protein when either *trp*LE1 or *trp*LE2, as compared to unimproved *trp*LE, is used in *E. coli* K12 MM294 to produce the fused gene product.

Claims

1. A shortened derivative of the *E. coli* tryptophan operon promoter-operator/leader/tryptophan E gene sequence from which sequences encoding the carboxy terminus of the leader peptide, the attenuator region of the leader sequence and the amino terminal region of the *trp*E gene have been deleted, which comprises a leader/tryptophan E gene sequence contained within about 428 to about 548 deoxyribonucleotide pairs.

2. The derivative of Claim 1 wherein the leader/tryptophan E gene sequence is contained within about 428 deoxyribonucleotide pairs.

3. The derivative of Claim 2 in which the gene is designated *trp*LE1 and in which the nucleotide sequence of the coding strand of the gene is

```
GA ATT CAC GCT GTG GTG TTA TGG TCG GTG GTC GCT AGG
GTG CCG ACG CGC ATC TCG ACT GCA CGG TGC ACC AAT GCT
TCT GGC GTC AGG CAG CCA ATC GGA AGC TGT GGT ATG GCT
GTG CAG GTC GTA TAA TCA CCG CAT AAT TCG AGT CGC TCA
AGG CGC ACT CCC GTT CCG GAT AAT GTT TTT TGC TCC GAC
ATC ATA ACG GTT CCG GCA AAT ATT CTG AAA TGA GCT GTT
GAC AAT TAA TCA TCG AAC TAG TTA ACT AGT ACG CAA GTT
CAC GTA AAA AGG GTA TCG ACA ATG AAA GCA ATT TTC GTA
CTG AAA GGT TCA CTG GAC AGA GAT CAT TCT GTT CCG CAG
TCG GAA GCC GAC GAA ACC CGT AAC AAA GCC CGC GCT GTA
CTG CGC GCT ATT GCC ACC GCG CAT CAT GCA CAG GAA TTC
```

wherein A is deoxyadenyl, G is deoxyguanyl, C is deoxycytidyl and T is thymidyl.

4. The derivative of Claim 1 wherein the leader/tryptophan E gene sequence is contained within about 548 deoxyribonucleotide pairs.

5. The derivative of Claim 3 in which the gene is designated *trp*LE2 and in which the nucleotide sequence of the coding strand of the gene is

```
GA ATT CAC GCT GTG GTG TTA TGG TCG GTG GTC GCT AGG
GTG CCG ACG CGC ATC TCG ACT GCA CGG TGC ACC AAT GCT
TCT GGC GTC AGG CAG CCA ATC GGA AGC TGT GGT ATG GCT
GTG CAG GTC GTA TAA TCA CCG CAT AAT TCG AGT CGC TCA
AGG CGC ACT CCC GTT CCG GAT AAT GTT TTT TGC TCC GAC
ATC ATA ACG GTT CCG GCA AAT ATT CTG AAA TGA GCT GTT
GAC AAT TAA TCA TCG AAC TAG TTA ACT AGT ACG CAA GTT
CAC GTA AAA AGG GTA TCG ACA ATG AAA GCA ATT TTC GTA
CTG AAA GGT TCA CTG GAC GGC AGC TAC GGC GGC GCG GTA
```

GGT TAT TTC ACC GCG CAT GGC GAT CTC GAC ACC TGC ATT
GTG ATC CGC TCG GCG CTG GTG GAA AAC GGT ATC GCC ACC
GTG CAA GCG GGT GCT GGT GTA GTC CTT GAT TCT GTT CCG
CAG TCG GAA GCC GAC GAA ACC CGT AAC AAA GCC CGC GCT
GTA CTG CGC GCT ATT GCC ACC GCG CAT CAT GCA CAG GAA
TTC

wherein A is deoxyadenyl, G is deoxyguanyl, C is deoxycytidyl and T is thymidyl.

6. A DNA sequence which comprises in the 5′ to 3′ direction the tryptophan promoter-operator/leader/tryptophan E gene sequence of any of Claims 1 to 5 immediately adjacent to and in transcriptional reading phase with an insulin-like growth factor II structural gene wherein the coding strand of the growth factor structural gene has the nucleotide sequence:

5' G CTTATCGACC GTCTGAAACT CTGTGCGGCG GCGAACTGGT
TGACACTCTG CAGTTCGTTT GCGGCGACCG TGGCTTCTAC
TTCTCTCGTC CGGCTTCTCG TGTTTCTAGA CGTTCTCGTG
GCATCGTTGA AGAATGCTGC TTCCGCTCTT GCGACCTGGC
TCTGCTGGAA ACTTACTGCG CTACTCCTGC TAAATCTGAA 3'

wherein A is deoxyadenyl, G is deoxyguanyl, C is deoxycytidyl and T is thymidyl.

7. A recombinant DNA expression vector comprising:

a) a tryptophan promoter-operator/leader/tryptophan E gene sequence contained within about 428 to about 548 nucleotides;

b) a structural gene encoding a heterologous polypeptide, positioned for expression and located downstream and in translational reading phase with (a); and

c) a replicon and selectable marker that provide for vector replication and allow for selection in a microorganism transformed by the vector.

8. The vector of Claim 7 which is a plasmid.

9. The vector of Claim 7 or 8 wherein the replicon and selectable marker allow for vector replication and selection in *E. coli*.

10. The vector of Claim 7, 8 or 9 wherein the selectable marker confers antibiotic resistance.

11. The vector of Claim 10 wherein the replicon and resistance-conferring selectable marker are from plasmid pBR322, pBR325, pBR324 or pKN402.

12. The vector of any of Claims 7 to 11 wherein the structural gene codes for insulin-like growth factor I, insulin-like growth factor II, bovine growth hormone, human growth hormone, human pre-growth hormone, porcine growth hormone, mammalian growth hormone, avian growth hormone, human insulin A chain, human insulin B chain, human proinsulin, human pre-proinsulin, interferon, urokinase, human tissue plasminogen activator, growth hormone releasing factor or interleukin II.

13. The vector of Claim 12 wherein the structural gene is positioned for expression immediately adjacent to, downstream of, and in translational reading phase with the leader/tryptophan E gene sequence of about 428 to about 548 deoxyribonucleotide pairs.

14. The vector of Claim 12 or 13 wherein the leader/tryptophan E gene sequence is the *trp*LE1 sequence of Claim 3.

15. The vector of Claim 12 or 13 wherein the leader/tryptophan E gene sequence is the *trp*LE2 sequence of Claim 5.

16. The vector of Claim 14 which is plasmid pCZ20 obtainable from *E. coli* K12 RV308/pCZ20 (NRRL B-15881) and shown in Figure 1.

17. The vector of Claim 15 which is plasmid pLEBGH2 obtainable from *E. coli* K12 RR1/pLEBGH2 (NRRL B-15882) and shown in Figure 2.

18. The vector of Claim 14 which is plasmid pLEBGH1 obtainable by ligating the ~5.5 kg *Eco*RI restriction fragment of plasmid pLEBGH2 and the ~0.43 kb *Eco*RI restriction fragment of plasmid pCZ20.

19. The plasmid obtainable by ligating the *Eco*RI-*Bam*HI digest of plasmid pBR322, the *trp*LE1-encoding ~0.43 kb *Eco*RI restriction fragment of plasmid pCZ20, and

(1) the ~0.425 kb insulin A chain structural gene-containing *Eco*RI-*Bam*HI restriction fragment of plasmid plA1, which is designated plALE1, or

(2) the ~0.453 kb insulin B chain structural gene-containing *Eco*RI-*Bam*HI restriction fragment of plasmid plB1, which is designated plBLE1, or

(3) the ~0.705 kb proinsulin structural gene-containing *Eco*RI-*Bam*HI restriction fragment of plasmid pHI7Δ4Δ1, which is designated pPILE1.

20. The plasmid obtainable by ligating the *Eco*RI-*Bam*HI digest of plasmid pBR322, the *trp*LE2-encoding ~0.55 kb *Eco*RI restriction fragment of plasmid PLEBGH2, and

(1) the ~0.425 kb insulin A chain structural gene-containing *Eco*RI-*Bam*HI restriction fragment of plasmid pIA1, which is designated pIALE2, or

(2) the ~0.453 kb insulin B chain structural gene-containing *Eco*RI-*Bam*HI restriction fragment of plasmid pIB1, which is designated pIBLE2, or

(3) the ~0.705 kb proinsulin structural gene-containing *Eco*RI-*Bam*HI restriction fragment of plasmid pHI7Δ4Δ1, which is designated pPILE2.

21. The plasmid obtainable by ligating the *Eco*RI-*Bam*HI digest of plasmid pBR322, a DNA sequence comprising the IGFII structural gene and having an *Eco*RI overlap at one end and a *Bam*HI overlap at the other end, and

(1) the *trp*LE1-encoding ~0.43 kb *Eco*RI restriction fragment of plasmid pCZ20, which is designated pIGF201 and shown in Figure 3, or

(2) the *trp*LE2-encoding ~0.55 kb *Eco*RI restriction fragment of plasmid pLEBGH2, which is designated pIGF202 and shown in Figure 4.

22. A transformed host cell which is *E. coli* R/R′ wherein R is K12, K12 RV308, K12 HB101, K12 C600R_k-M_k-, K12 RR1, or K12 MM294 and where R′ is the vector of any of Claims 7 to 13 or the plasmid of Claim 19, 20 or 21.

Patentansprüche

1. Verkürztes Derivat der E. coli-Tryptophan-Operon-Promotor-Operator/Leader/Tryptophan E-Gensequenz, aus der Sequenzen, die für das Carboxyende des Leaderpeptids, den Attenuatorbereich der Leadersequenz und den aminoendständigen Bereich des trpE-Gens kodieren, deletiert worden sind, umfassend eine Leader/Tryptophan E-Gensequenz, die innerhalb von etwa 428 — etwa 548 Desoxyribonucleotid-Paaren enthalten ist.

2. Derivat nach Anspruch 1, wobei die Leader/Tryptophan E-Gensequenz innerhalb von etwa 428 Desoxyribonucleotid-Paaren enthalten ist.

3. Derivat nach Anspruch 2, bei dem das Gen die Bezeichnung trpLE1 aufweist und bei dem folgende Nucleotidsequenz des kodierenden Strangs des Gens vorliegt:

```
GA ATT CAC GCT GTG GTG TTA TGG TCG GTG GTC GCT AGG
GTG CCG ACG CGC ATC TCG ACT GCA CGG TGC ACC AAT GCT
TCT GGC GTC AGG CAG CCA ATC GGA AGC TGT GGT ATG GCT
GTG CAG GTC GTA TAA TCA CCG CAT AAT TCG AGT CGC TCA
AGG CGC ACT CCC GTT CCG GAT AAT GTT TTT TGC TCC GAC
ATC ATA ACG GTT CCG GCA AAT ATT CTG AAA TGA GCT GTT
GAC AAT TAA TCA TCG AAC TAG TTA ACT AGT ACG CAA GTT
CAC GTA AAA AGG GTA TCG ACA ATG AAA GCA ATT TTC GTA
CTG AAA GGT TCA CTG GAC AGA GAT CAT TCT GTT CCG CAG
TCG GAA GCC GAC GAA ACC CGT AAC AAA GCC CGC GCT GTA
CTG CGC GCT ATT GCC ACC GCG CAT CAT GCA CAG GAA TTC
```

worin A Desoxyadenyl bedetuet, G Desoxyguanyl bedeutet, C Desoxycytidyl bedeutet und T Thymidyl bedeutet.

4. Derivat nach Anspruch 1, wobei die Leader/Tryptophan E-Gensequenz innerhalb von etwa 548 Desoxyribonucleotid-Paaren enthalten ist.

5. Derivat nach Anspruch 3, bei dem das Gen die Bezeichnung trpLE2 aufweist und bei dem die folgende Nucleotidsequenz des kodierenden Strangs des Gens vorliegt:

```
GA ATT CAC GCT GTG GTG TTA TGG TCG GTG GTC GCT AGG
GTG CCG ACG CGC ATC TCG ACT GCA CGG TGC ACC AAT GCT
TCT GGC GTC AGG CAG CCA ATC GGA AGC TGT GGT ATG GCT
GTG CAG GTC GTA TAA TCA CCG CAT AAT TCG AGT CGC TCA
```

AGG CGC ACT CCC GTT CCG GAT AAT GTT TTT TGC TCC GAC
ATC ATA ACG GTT CCG GCA AAT ATT CTG AAA TGA GCT GTT
GAC AAT TAA TCA TCG AAC TAG TTA ACT AGT ACG CAA GTT
CAC GTA AAA AGG GTA TCG ACA ATG AAA GCA ATT TTC GTA
CTG AAA GGT TCA CTG GAC GGC AGC TAC GGC GGC GCG GTA
GGT TAT TTC ACC GCG CAT GGC GAT CTC GAC ACC TGC ATT
GTG ATC CGC TCG GCG CTG GTG GAA AAC GGT ATC GCC ACC
GTG CAA GCG GGT GCT GGT GTA GTC CTT GAT TCT GTT CCG
CAG TCG GAA GCC GAC GAA ACC CGT AAC AAA GCC CGC GCT
GTA CTG CGC GCT ATT GCC ACC GCG CAT CAT GCA CAG GAA
TTC

wobei A Desoxyadenyl bedeutet, G Desoxyguanyl bedeutet, C Desoxycytidyl bedeutet und T Thymidyl bedeutet.

6. DNA-Sequenz, die in der 5′ bis 3′-Richtung die Tryptophan-Promotor-Operator/Leader/Tryptophan E-Gensequenz nach einem der Ansprüche 1 bis 5 in unmittelbarer Nachbarschaft und in transskriptionaler Lesephase mit einem insulinähnlichen Wachstumsfaktor II-Strukturgen enthält, wobei der kodierende Strang des Wachstumsfaktors-Strukturgen folgende Nucleotidsequenz aufweist:

5′ G CTTATCGACC GTCTGAAACT CTGTGCGGCG GCGAACTGGT
TGACACTCTG CAGTTCGTTT GCGGCGACCG TGGCTTCTAC
TTCTCTCGTC CGGCTTCTCG TGTTTCTAGA CGTTCTCGTG
GCATCGTTGA AGAATGCTGC TTCCGCTCTT GCGACCTGGC
TCTGCTGGAA ACTTACTGCG CTACTCCTGC TAAATCTGAA 3′

wobei A Desoxyadenyl bedeutet, G Desoxyguanyl bedeutet und C Desoxycytidyl bedeutet und T Thymidyl bedeutet.

7. Rekombinanter DNA-Expressionsvektor, umfassend:

a) eine Tryptophan-Promotor-Operator/Leader/Tryptophan E-Gensequenz, die innerhalb von etwa 428 bis etwa 548 Nucleotiden enthalten ist;

b) ein für ein heterologes Polypeptid kodierendes Strukturgen, das für die Expression positioniert ist und sich stromabwärts und in translationaler Lesephase zu (a) befindet,

(c) ein Replikon und einen selektierbaren Marker, der eine Vektorreplikation gewährleistet und eine Selektion in einem durch den Vektor transformierten Mikroorganismus ermöglicht.

8. Vektor nach Anspruch 7, bei dem es sich um ein Plasmid handelt.

9. Vektor nach Anspruch 7 oder 8, wobei das Replikon und der selektierbare Marker eine Replikation und Selektion des Vektors in E. coli ermöglichen.

10. Vektor nach Anspruch 7, 8 oder 9, wobei der selektierbare Marker antibiotische Resistenz verleiht.

11. Vektor nach Anspruch 10, wobei das Replikon und der resistenzverleihende selektierbare Marker aus Plasmid pBR322, pBR325, pBR324 oder pKN402 stammen.

12. Vektor nach einem der Ansprüche 7 bis 11, wobei das Strukturgen für insulinähnlichen Wachstumsfaktor I, insulinähnlichem Wachstumsfaktor II, Rinderwachstumshormon, humanes Wachstumshormon, humanes Präwachstumshormon, Schweinewachstumshormon, Säugetierwachstumshormon, Vogelwachstumshormon, humane Insulin A-Kette, humane Insulin B-Kette, humanes Proinsulin, humanes Präproinsulin, Interferon, Urokinase, humanen Gewebe-Plasminogen-Aktivator, Wachstumshormon-Freisetzungsfaktor oder Interleukin II kodiert.

13. Vektor nach Anspruch 12, wobei das Strukturgen sich für die Expression unmittelbar neben, stromabwärts zu und in translationaler Lesephase mit der Leader/Tryptophan E-Gensequenz von etwa 428 bis etwa 548 Desoxyribonucleotid-Paaren befindet.

14. Vektor nach Anspruch 12 oder 13, wobei es sich bei der Leader/Tryptophan E-Gensequenz um die trpLE1-Sequenz von Anspruch 3 handelt.

15. Vektor nach Anspruch 12 oder 13, wobei es sich bei der Leader/Tryptophan E-Gensequenz um die trpLE2-Sequenz von Anspruch 5 handelt.

16. Vektor nach Anspruch 14, bei dem es sich um das Plasmid pCZ20 handelt, das aus E. coli K12 RV308/pCZ20 (NRRL B-15881) erhältlich ist und in Fig. 1 gezeigt ist.

17. Vector nach Anspruch 15, bei dem es sich um das Plasmid pLEBGH2 handelt, das aus E. coli K12 RR1/pLEBGH2 (NRRL B-15882) erhältlich ist und in Fig. 2 gezeigt ist.

18. Vektor nach Anspruch 14, bei dem es sich um das Plasmid pLEBGH1 handelt, das durch Verknüpfung des ~5,5 kb-EcoRI-Restriktionsfragments von Plasmid pLEBGH2 und des ~0,43 kb-EcoRI-Restriktionsfragments von Plasmid pCZ20 erhältlich ist.

19. Plasmid, erhältlich durch Verknüpfen des EcoRI-BamHI-Verdauungsprodukts von Plasmid pBR322, des trpLE1-kodierenden ~0,43 kb-EcoRI-Restriktionsfragments von Plasmid pCZ20 und

(1) des das ~0,425 kb-Insulin A-Kette-Strukturgen enthaltenden EcoRI-BamHI-Restriktionsfragments von Plasmid pIA1, das als pIALE1 bezeichnet wird, oder

(2) des das ~0,453 kb-Insulin-B-Kette-Strukturgens enthaltenden EcoRI-BamHI-Restriktionsfragments von Plasmid pIB1, das als pIBLE1 bezeichnet wird oder

(3) des das ~0,705 kb-Proinsulin-Strukturgen enthaltenden EcoRI-BamHI-Restriktionsfragments von Plasmid pHI7Δ4Δ1, das mit pPILE1 bezeichnet wird.

20. Plasmid, erhältlich durch Verknüpfen des EcoRI-BamHI-Verdauungsprodukts von Plasmid pBR322, des trpLE2-kodierenden ~0,55 kb-EcoRI-Restriktionsfragments von Plasmid PLEBGH2, und

(1) des das ~0,425 kb-Insulin A-Kette-Strukturgen enthaltenden EcoRI-BamHI-Restriktionsfragments von Plasmid pIA1, das mit pIALE2 bezeichnet wird, oder

(2) des das ~0,453 kb-Insulin-B-Kette-Strukturgens enthaltenden EcoRI-BamHI-Restriktionsfragments von Plasmid pIB1, das mit pIBLE2 bezeichnet wird oder

(3) des das ~0,705 kb-Proinsulin-Strukturgen enthaltenden EcoRI-BamHI-Restriktionsfragments von Plasmid pHI7Δ4Δ1, das mit pPILE2 bezeichnet wird.

21. Plasmid erhältlich durch Verknüpfen des EcoRI-BamHI-Verdauungsprodukts des Plasmids pBR322, einer DNA-Sequenz mit einem Gehalt an dem IGFII-Strukturgen mit einer EcoRI-Überlappung an einem Ende und einer BamHI-Überlappung am anderen Ende und

(1) des trpLE1-kodierenden ~0,43 kb EcoRI-Restriktionsfragments von Plasmid pCZ20, das mit pIGF201 bezeichnet wird und in Fig. 3 gezeigt ist oder

(2) des trpLE2-kodierenden ~0,55 kb-EcoRI-Restriktionsfragments von Plasmid pLEBGH2, das mit pIGF202 bezeichnet wird und in Fig. 4 gezeigt ist.

22. Transformierte Wirtszelle, bei der es sich um E. coli R/R' handelt, wobei R K12, K12 RV308, K12 HB101, K12 C600R_k-M_k-, K12 RR1 oder K12 MM294 ist und wobei R' den Vektor nach einem der Ansprüche 7 bis 13 oder das Plasmid von Anspruch 19, 20 oder 21 bedeutet.

Revendications

1. Dérivé raccourci de la séquence de gène promoteur-opérateur directeur/tryptophane E de l'opéron tryptophane de E. coli dont les séquences codant pour l'extrémité carboxyle du peptide directeur, la région d'atténuateur de la séquence directrice et la région terminale amino du gène trpE ont été supprimées, qui comprend une séquence de gène directeur/tryptophane E contenue dans environ 428 à environ 548 paires de désoxyribonucléotides.

2. Dérivé selon la revendication 1, dans lequel la séquence du gène dirigeant/tryptophane E est contenue dans environ 428 paires de désoxyribonucléotides.

3. Dérivé selon la revendication 2, dans lequel le gène est désigné trpLE1 et dans lequel la séquence de nucléotide du brin codant du gène est

```
GA ATT CAC GCT GTG GTG TTA TGG TCG GTG GTC GCT AGG
GTG CCG ACG CGC ATC TCG ACT GCA CGG TGC ACC AAT GCT
TCT GGC GTC AGG CAG CCA ATC GGA AGC TGT GGT ATG GCT
GTG CAG GTC GTA TAA TCA CCG CAT AAT TCG AGT CGC TCA
AGG CGC ACT CCC GTT CCG GAT AAT GTT TTT TGC TCC GAC
ATC ATA ACG GTT CCG GCA AAT ATT CTG AAA TGA GCT GTT
GAC AAT TAA TCA TCG AAC TAG TTA ACT AGT ACG CAA GTT
CAC GTA AAA AGG GTA TCG ACA ATG AAA GCA ATT TTC GTA
CTG AAA GGT TCA CTG GAC AGA GAT CAT TCT GTT CCG CAG
TCG GAA GCC GAC GAA ACC CGT AAC AAA GCC CGC GCT GTA
CTG CGC GCT ATT GCC ACC GCG CAT CAT GCA CAG GAA TTC
```

dans laquelle A représente un reste désoxyadényle, G un reste désoxyguanyle, C un reste désoxycytidyle et T un reste thymidyle.

4. Dérivé selon la revendication 1, dans lequel la séquence de gène directeur/tryptophane E est contenue dans environ 548 paires de désoxyribonucléotides.

5. Dérivé selon la revendication 3, dans lequel le gène est désigné trpLE2 et dans lequel la séquence de nucléotide du brin codant du gène est

GA ATT CAC GCT GTG GTG TTA TGG TCG GTG GTC GCT AGG
GTG CCG ACG CGC ATC TCG ACT GCA CGG TGC ACC AAT GCT
TCT GGC GTC AGG CAG CCA ATC GGA AGC TGT GGT ATG GCT
GTG CAG GTC GTA TAA TCA CCG CAT AAT TCG AGT CGC TCA
AGG CGC ACT CCC GTT CCG GAT AAT GTT TTT TGC TCC GAC
ATC ATA ACG GTT CCG GCA AAT ATT CTG AAA TGA GCT GTT
GAC AAT TAA TCA TCG AAC TAG TTA ACT AGT ACG CAA GTT
CAC GTA AAA AGG GTA TCG ACA ATG AAA GCA ATT TTC GTA
CTG AAA GGT TCA CTG GAC GGC AGC TAC GGC GGC GCG GTA
GGT TAT TTC ACC GCG CAT GGC GAT CTC GAC ACC TGC ATT
GTG ATC CGC TCG GCG CTG GTG GAA AAC GGT ATC GCC ACC
GTG CAA GCG GGT GCT GGT GTA GTC CTT GAT TCT GTT CCG
CAG TCG GAA GCC GAC GAA ACC CGT AAC AAA GCC CGC GCT
GTA CTG CGC GCT ATT GCC ACC GCG CAT CAT GCA CAG GAA
TTC

dans laquelle A représente un reste désoxyadényle, G un reste désoxyguanyle, C un reste désoxycytidyle et T un reste thymidyle.

6. Séquence d'ADN comprenant, dans le sensn 5' à 3', la séquence de gène tryptophane promoteur/opérateur/directeur/tryptophane E selon l'une des revendications 1 à 5, immédiatement voisine de et en phase de lecture transcriptionnelle avec un gène structurel de facteur de croissance II analogue à l'insuline, dans lequel le brin codant du gène structurel de facteur de croissance comporte la séquence de nucléotides

5' G CTTATCGACC GTCTGAAACT CTGTGCGGCG GCGAACTGGT
TGACACTCTG CAGTTCGTTT GCGGCGACCG TGGCTTCTAC
TTCTCTCGTC CGGCTTCTCG TGTTTCTAGA CGTTCTCGTG
GCATCGTTGA AGAATGCTGC TTCCGCTCTT GCGACCTGGC
TCTGCTGGAA ACTTACTGCG CTACTCCTGC TAAATCTGAA 3'

dans laquelle A représente une désoxyadényle, G une désoxyguanyle, C une désoxycytidyne et T une thymidyne.

7. Vecteur d'expression de l'ADN recombinant comprenant:

a) la séquence de gène tryptophane promoteur/opérateur/directeur/tryptophane E contenue dans environ 428 à environ 548 nucléotides;

b) un gène structurel codant pour un polypeptide hétérologue disposé pour l'expression et situé en amont et en phase de lecture translationnelle avec (a); et

c) un réplicon et un marqueur sélectionnable qui assurent la réplication du vecteur et permettent la sélection dans un micro-organisme transformé par le vecteur.

8. Vecteur selon la revendication 7 qui est un plasmide.

9. Vecteur selon la revendication 7 ou 8, dans lequel le réplicon et le marqueur sélectionnable permettent la réplication du vecteur et la sélection dans E. coli.

10. Vecteur selon la revendication 7, 8 ou 9, dans lequel le marqueur sélectionnable confère une résistance aux antibiotiques.

11. Vecteur selon la revendication 10, dans lequel le réplicon et le marqueur sélectionnable conférant une résistance proviennent du plasmide pBR322, pBR325, pBR324 et pKN402.

12. Vecteur selon l'une des revendications 7 à 11, dans lequel les gènes structurels codent pour le facteur de croissance I analogue à l'insuline, le facteur de croissance II analogue à l'insuline, l'hormone de

croissance bovine, l'hormone de croissance humaine, l'hormone de précroissance humaine, l'hormone de croissance porcine, l'hormone de croissance des mammifères, l'hormone de croissance des oiseaux, la chaîne de l'insuline A humaine, la chaîne de l'insuline B humaine, la pro-insuline humaine, la pré-pro-insuline humaine, l'interféron, l'urokinase, l'activateur du plasminogène des tissus humains, le facteur dégageant l'hormone de croissance ou l'interleukine II.

13. Vecteur selon la revendication 12, dans lequel le gène structurel est disposé pour l'expression immédiatement au voisinage et en aval de et en phase de lecture transactionnelle avec la séquence de gène directeur/tryptophane E d'environ 428 à environ 548 paires de désoxyribonucléotides.

14. Vecteur selon la revendication 12 ou 13, dans lequel la séquence de gène directeur/tryptophane E est la séquence trpLE1 de la revendication 3.

15. Vecteur selon la revendication 12 ou 13, dans lequel la séquence de gène directeur/tryptophane E est la séquence trpLE2 de la revendication 5.

16. Vecteur selon la revendication 14, qui est le plasmide pCZ20 pouvant être obtenu au départ de K12 RV308/pCZ20 de E. coli (NRRL B-15881), comme le montre la figure 1.

17. Vecteur selon la revendication 15, qui est le plasmide pLEBGH2 pouvant être obtenu au départ de K12 RR1/pLEBGH2 de E. coli (NRRL B-15882), comme le montre la figure 2.

18. Vecteur selon la revendication 14, qui est le plasmide pLEBGH1 pouvant être obtenu en ligant le fragment de restriction d'environ 5,5 kb de EcoRI du plasmide pLEBGH2 et le fragment de restriction d'environ 0,43 kb de EcoRI du plasmide pCZ20.

19. Plasmide pouvant être obtenu en ligant le produit de digestion par EcoRI et par BamHI du plasmide pBR322, le fragment de restriction d'environ 0,43 kb digéré par EcoRI en encodant trpLE1 du plasmide pCZ20, et

(1) le fragment de restriction d'environ 0,425 kb digéré par EcoRI et BamHI contenant le gène structurel de la chaîne de l'insuline A du plasmide pIA1, qui est désigné pIALE1, ou

(2) le fragment de restriction d'environ 0,453 kb digéré par EcoRI et BamHI contenant le gène structurel de la chaîne de l'insuline B du plasmide pIB1 qui est désigné pIBLE1, ou

(3) le fragment de restriction d'environ 0,705 kb digéré par EcoRI et BamHI contenant le gène structurel de la pro-insuline du plasmide pHI7Δ4Δ1, qui est désigné pPILE1.

20. Plasmide pouvant être obtenu en ligant le produit de la digestion par EcoRI et BamHI du plasmide pBR322, le fragment de restriction d'environ 0,55 kb digéré par EcoRI et comprenant le trpLE2 du plasmide PLEBGH2, et

(1) le fragment de restriction d'environ 0,425 kb digéré par EcoRI et BamHI contenant le gène structurel de la chaîne de l'insuline A du plasmide pIA1, qui est désigné pIALE2, ou

(2) le fragment de restriction d'environ 0,453 kb digéré par EcoRI et BamHI contenant le gène structurel de la chaîne de l'insuline B du plasmide pIB1 qui est désigné pIBLE2, ou

(3) le fragment de restriction d'environ 0,705 kb digéré par EcoRI et BamHI contenant le gène structurel de la pro-insuline du plasmide pHI7Δ4Δ1, qui est désigné pPILE2.

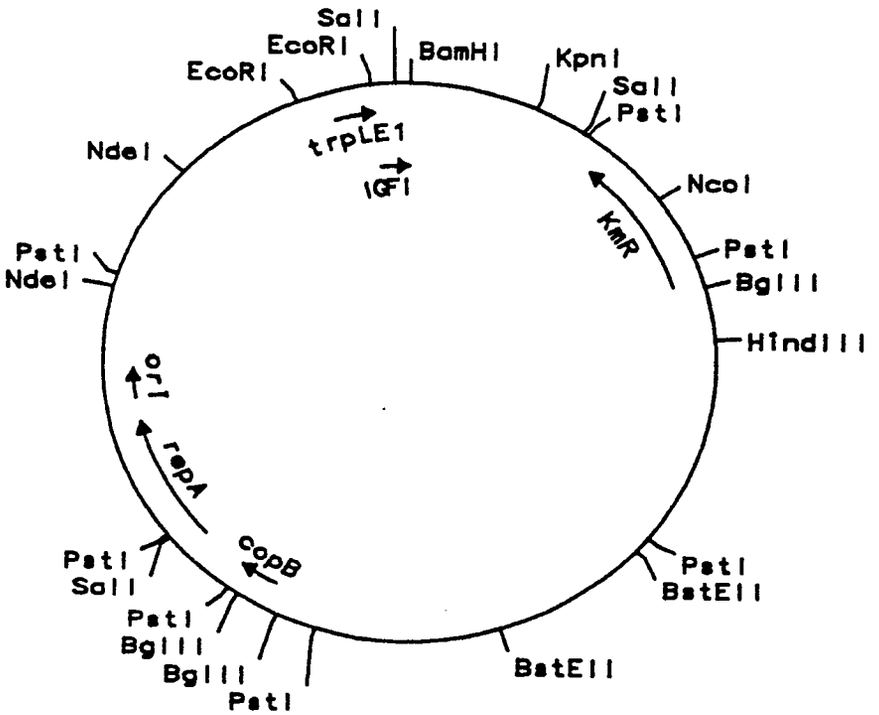
21. Plasmide pouvant être obtenu en ligant le produit de digestion par EcoRI et BamHI du plasmide pBR322, une séquence d'ADN comprenant le gène structurel IGFII et ayant un recouvrement EcoRI à une extrémité et un recouvrement BamHI à l'autre extrémité, et

(1) le fragment de réstriction d'environ 0,43 kb digéré par EcoRI et codant pour trpLE1 du plasmide pCZ20 qui est désigné pIGF201 et est représenté à la figure 3, ou

(2) le fragment de restriction d'environ 0,55 kb digéré par EcoRI codant pour trpLE2 du plasmide pLEBGH2, qui est désigné par pIGF202 et est représenté à la figure 4.

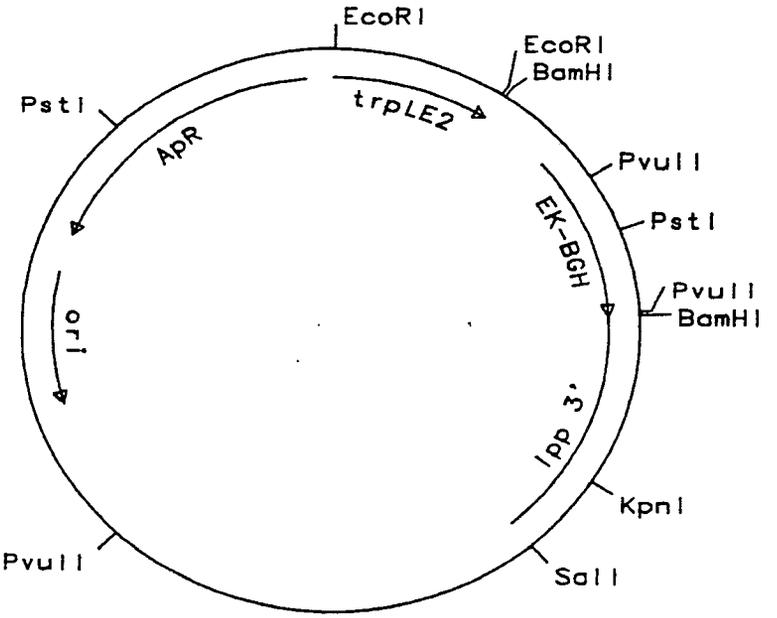
22. Cellule hôte transformée qui est E. coli R/R', dans lequel R représente K12, K12 RV308, K12 HB101, K12 C600R_k-M_k-, K12 RR1 ou K12 MM294 et dans lequel R' est le vecteur de l'une des revendications 7 à 13 ou le plasmide de la revendication 19, 20 ou 21.

FIG.1



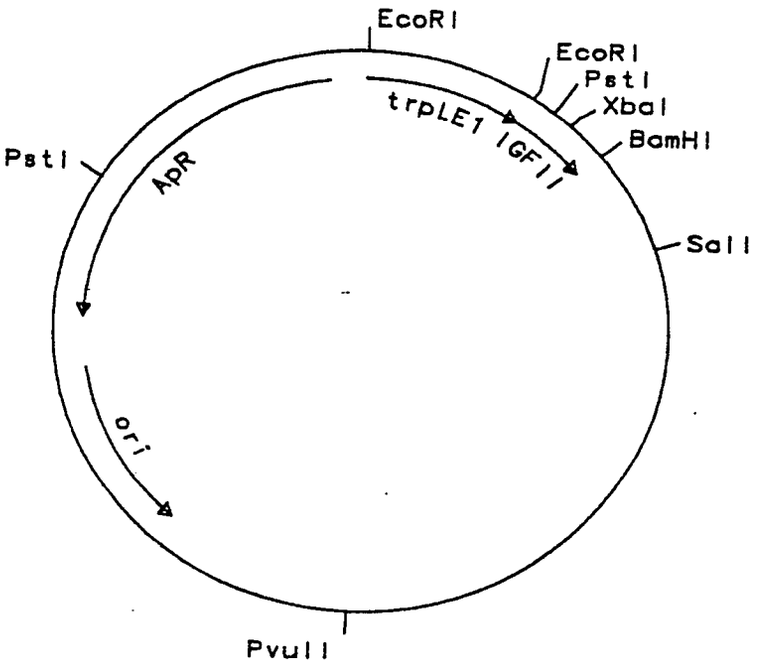
Restriction Site and Function Map of Plasmid pCZ20 (10.7kb)

FIG.2



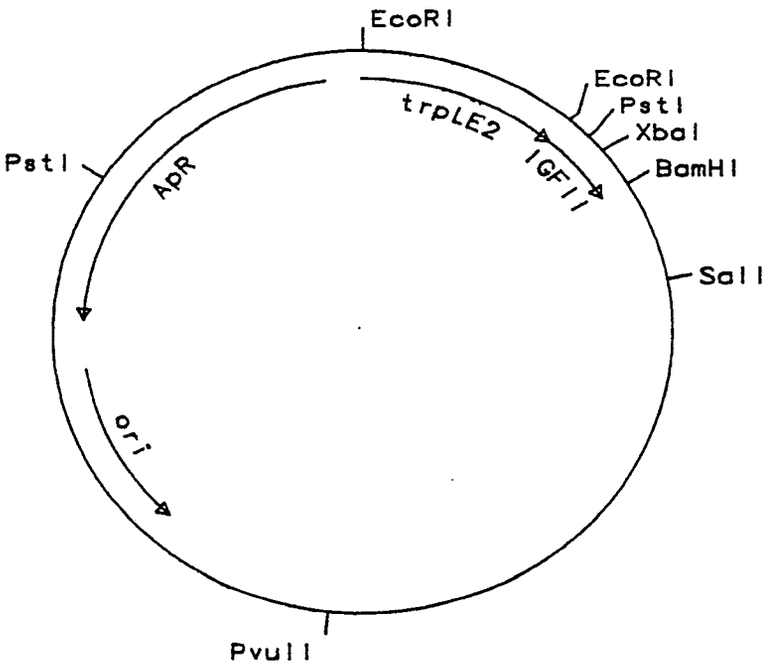
Restriction Site and Function Map of
Plasmid pLEBGH2
(6.0kb)

FIG.3



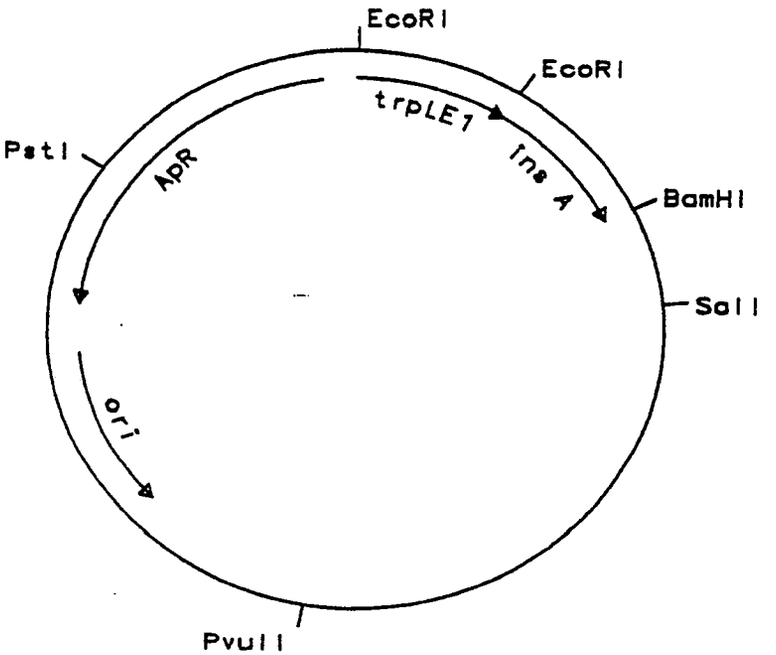
Restriction Site and Function Map of Plasmid pIGF201 (4.6kb)

FIG. 4



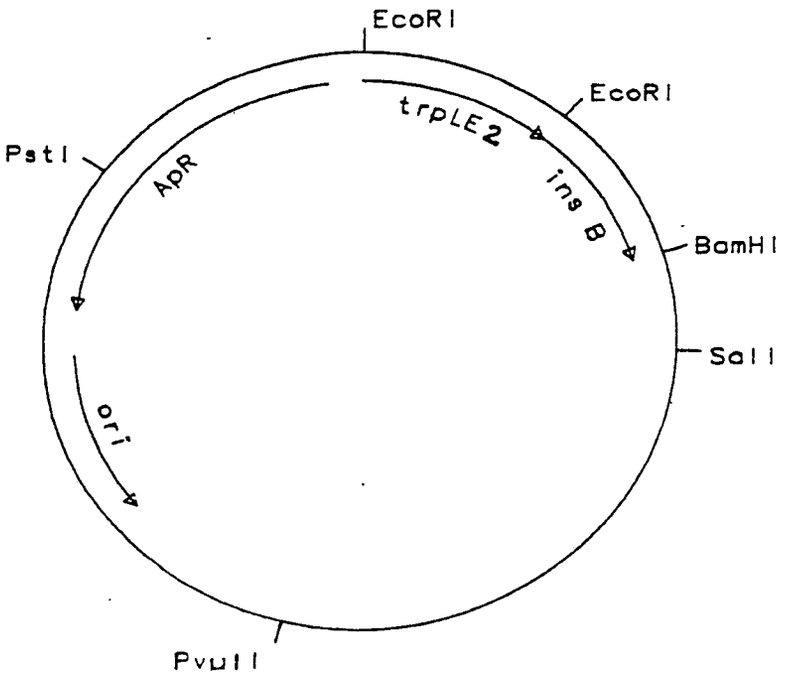
Restriction Site and Function Map of
Plasmid pIGF202
(4.8kb)

FIG.5



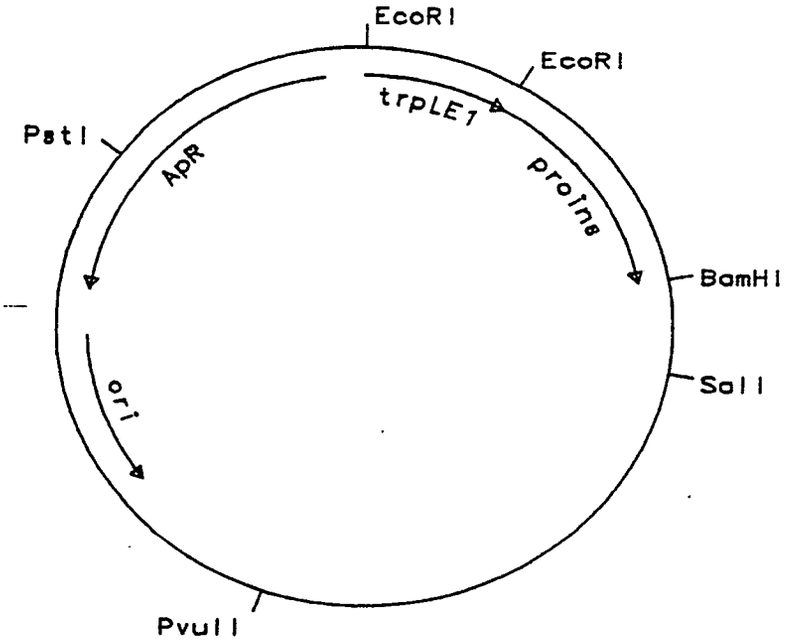
Restriction Site and Function Map of Plasmid pIALE1 (4.8kb)

FIG.6



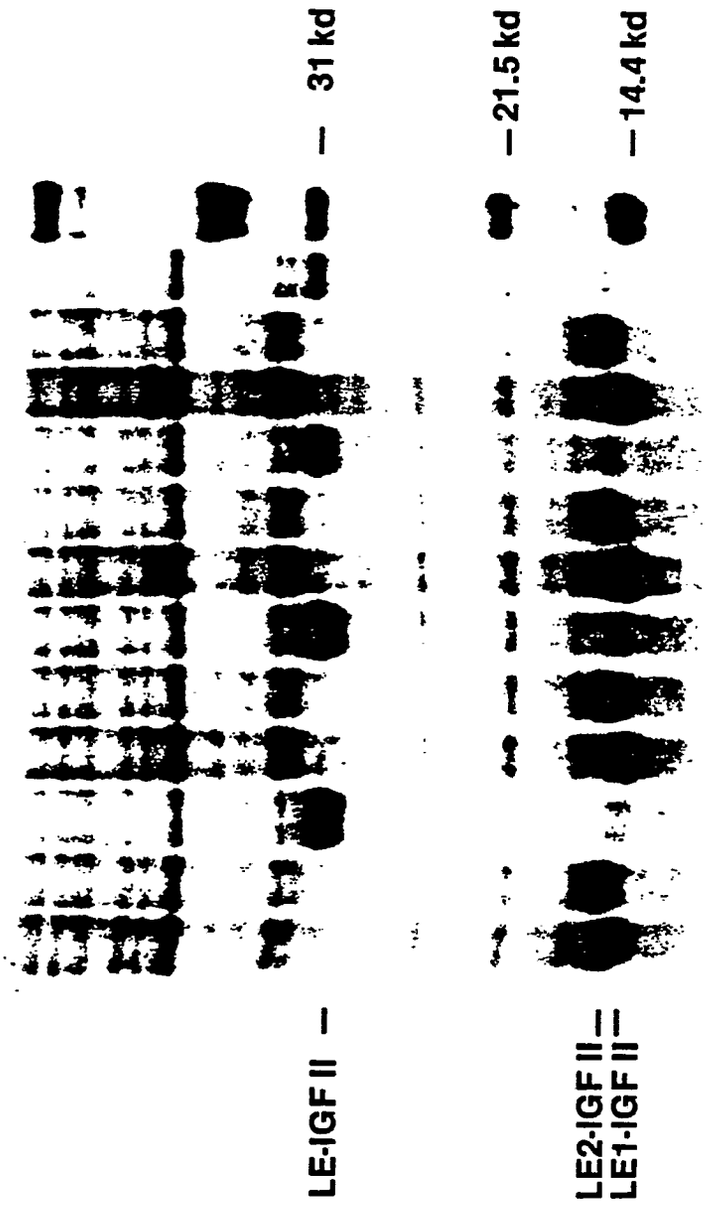
Restriction Site and Function Map of Plasmid pIBLE2 (5.0kb)

FIG.7



Restriction Site and Function Map of Plasmid pPILE1 (5.1kb)

FIG.8


1 2 3 4 5 6 7 8 9 10 11 12 13
31 kd
21.5 kd
14.4 kd
LE-IGF II
LE2-IGF II
LE1-IGF II